



Analytic or Impressionistic Evaluation of EFL Jordanian Students'

Composition Performance?

Dr. Khalaf Al-Makhzoomi

Department of Curriculum and Instruction, Faculty of Education

Yarmouk University, Jordan

Dr. Saleh Freihat

English Department, Faculty of Arts

Isra University, Jordan

Abstract

Evaluating composition performance among English as a foreign language (EFL) students has become increasingly difficult because of the nature of the writing skill itself. This study discusses how EFL Jordanian teachers can use its results to help improve their students' writing ability. Specifically, it aims at finding a measure for EFL composition which would increase its grading reliability and improve its teaching to EFL Jordanian students. The researchers compared the analytic composition grading method with the impressionistic one. They graded a sample of 90 EFL Jordanian students' composition scripts impressionistically and then analytically using the analysis of variance to estimate the reliability for the scripts' graders on the 2 methods. The study showed that the analytic method has pedagogical advantages over the impressionistic one in that it lays the foundation of the relevant elements of good writing and when the writing teachers calculate the marks in the marking scheme for their students, they can discover some strength and see where their efforts are successful, where their students need special attention

and where to locate areas for improvement. Further pieces of research on EFL writing are recommended to document and analyse the practical pedagogical knowledge of highly experienced EFL Jordanian composition teachers.

Keywords: analytic / impressionistic evaluation; composition performance; EFL writing; EFL writing instruction.

Introduction

The evaluation of composition performance among English as a foreign/second language EFL/ESL students has become increasingly difficult because of the complex nature of the writing skill itself. Careful specification of a free writing grading instrument / tool that increases both grader's reliability as well as his/her objectivity is exactly what researchers have long been investigating for the sake of evaluating / measuring of EFL / ESL students' writing proficiency/ability. More specifically, pieces of research on this particular area are badly needed to facilitate, improve and make EFL/ESL teachers aware of and capable of grading and or evaluating their students' free writing tasks written in English.

Literature Review

The attempts to improve the grading of non-native students' free writing performance in an English as a foreign/second language (EFL/ESL) program are among the main areas various researchers handled. Kaczmarek (1980, 151-170) investigated the essay, i.e., English composition as a test of language proficiency and proceeded to compare the most debated two methods of evaluating the essay, i.e., *subjective vs objective methods*. The researcher found out that *subjective methods* worked about as well as *objective grading techniques*, and interestingly enough both were found to be strongly correlated with each other and with other measures of English as a Second Language (ESL) proficiency. Heaton (1982) indicated that graders/scorers meet the use of essays as a measure of language proficiency with opposition because of the existence of inter – rater reliability, i.e., two graders may give two different grades to the same essay. This point represented the problem which belonged to the assessment of free writing compositions. Heaton mentioned two methods of grading free writing which graders/scorers

use less frequently than the *analytic method*. The *impression method* of grading writing entails one or more graders awarding a single grade (multiple grading), based on the total impression of the composition topic as a whole. This method is generally more tiring and faster than the *analytic* and *mechanical grading*. The *mechanical accuracy (error – count) method* is the least valid of all methods of grading free writing and is not recommended simply because it is based on counting and deducing the number of errors from a given score. Since researchers can reach no decision about the relative importance of most errors, the whole scheme is highly subjective. What is more, this method ignores the real purpose of writing and focuses only on the negative aspects of a writing task, placing the student in such a position that he/she cannot write for fear of making mistakes. Homburg (1984, 87-103) investigated the process of holistic grading of compositions written by students of English as a Second Language (ESL). Although EFL/ESL writing instructors evaluate EFL/ESL writing proficiency in intensive English programs through the use of holistic grading of actual students' writing samples (probably because it seems that there is no better test of writing proficiency than an actual writing sample), this procedure is not without problems. Of the major problems, reliability- both inter – rater and intra – rater – is the most troublesome. First, different composition graders often assign the same composition to different grading categories, and this may affect inter – rater reliability. Second, the same composition grader may well assign the same composition to different grading categories at different times and this may affect intra – rater reliability. Although these are significant problems, it nonetheless appears that many ESL professionals consider the holistic grading of ESL compositions to be valid, as evidenced by its widespread use for the evaluation of ESL students' writing proficiency. Carroll and Hall (1985) indicated that there is a general agreement among teachers of English as a Foreign / Second Language (EFL/ESL) teachers, specialists and test designers as to the difficulty of grading a piece of free writing objectively. As far as methods of grading composition are concerned, Carroll and Hall seemed to support the idea that a reliable assessment of free, open-ended and creative writing/ composition is possible by means of careful co-ordination. Gannon (1985) supported the reliability of the *analytic grading method* in grading free writing tasks simply because many segments of free writing procedures lend themselves to objective testing such as *lexis, syntax, spelling and punctuation*. Brown and Baily (1987) pointed out that the careful specification of the analytical grading instrument can increase grader reliability, another term for the objectivity of the *analytic method*.

According to Heaton (1988), attention must be paid to the following writing skills: language use (*the ability to write correct sentences*), mechanical skills (*the ability to spell and punctuate*), treatment of content (*the ability to think creatively, develop thoughts and exclude irrelevant information*), stylistic skills (*the ability to manipulate sentences and paragraphs and use language effectively*), and judgment skills (*the ability to write in an appropriate manner for a particular purpose with a particular audience in mind*). Regarding composition – type tests, 3 researchers introduced, employed and advocated 3 major methods of grading: *analytic grading*

(Heaton, 1988), *holistic grading* (Kammeen, 1989, 162-170) and *frequency – count grading* (Wilkinson, 1989, 67-70). Among these three measures of evaluating compositions, the *frequency – count grading* is the most objective one and it is exactly what researchers have long been investigating. Wilkinson (1989) asserted that there are many elements graders consider in measuring the writing ability. These elements include *form, content, grammar, vocabulary, mechanics (including spelling and punctuation), handwriting, accuracy, style, diction, relevance, originality, elaboration, layout, coherence, unity, cohesion, organization and logic*.

Hackeling (1991, 13-28) conducted a study on error analysis of English compositions written by Japanese students to determine which sentence – level errors they make most frequently and of those errors which are the most serious. The majority of the surveys ranked the errors in the *tense area* as the least serious / most acceptable. Graders ranked *singular / plural errors* on the somewhat serious level. They also ranked *word form errors and word order errors* as the least acceptable / most annoying. For free writing analysis, EL-Kilabi (1996, 9 -26) conducted a study on 77 third – year subjects (males and females) from the English Department at Al-Turath University College, Baghdad, Iraq to establish the objectivity of the *analytic method* as a reliable instrument of grading free writing tasks. The researcher achieved this by using objective questions (error – recognition and error – correction questions) by establishing their relationship to essay writing questions scored using the *analytic method*, and then by deciding the magnitude of relationship between test questions and essay writing questions. The validity of the two writing tests was established by means of item analysis, and their reliability by means of a parallel form test (the Cloze Procedure test). The researcher found the 2 tests reliable. Farhady and Farzanehnejad (1996, 298-307) conducted a study to devise an objective measure for evaluating writing as a thinking process. They called it the "Measure of Cohesion (MC)". The results of the study indicated that graders/scores can compute such a measure in the following way:

$$\text{Measure of Cohesion (MC)} = \frac{\text{Number of Cohesive Devices}}{\text{Number of Words}}$$

According to Pennington, Costa, Shing, Hirose and Niedzielski (1997, 120-430), it would be a challenge for researchers to distinguish precisely between writing courses that adopt a specific-purpose or general- purpose orientation, given that these categories tend to overlap in many instances, particularly in academic contexts. A further limitation is that interview data present

information on what writing instructors may ideally want to be doing which may differ from what they do in practice when teaching writing classes. Cumming (1998 ,61-78) designed a piece of research to document and analyse the practical pedagogical knowledge of people who are highly experienced in teaching English as a Second / Foreign Language (ESL/EFL) written composition. Cumming and Riazi (2000, 55-71) believed that documenting how experienced teachers conceptualize their thinking about ESL/EFL writing instruction is a necessary step towards developing empirical models of second language writing instruction. Cumming (2001: 207 – 224) interviewed 48 highly experienced instructors of ESL/EFL composition about their usual practices for writing assessment in courses in universities or immigrant settlement programs. Although the instructors tended to conceptualize ESL/EFL writing instruction in a common way overall, the researcher was surprised to find how composition instructors' conceptualizations of student assessment varied depending on whether the writing courses they taught were defined in reference to general or specific purposes for learning English. Bacha (2001) carried out a study on a stratified random sample of final exam essays written by L1 Arabic non-native students of English attending the Freshman English I course in the English as a Foreign Language (EFL) program at the American University of Beirut. Specifically, this study aimed at finding out what *holistic essay writing* can tell EFL/ESL teachers and what general lessons can be drawn for the evaluation of writing. The study results indicated that writing instructors should pay more attention to the language and vocabulary aspects of students' essay writing and a combination

of holistic and analytic evaluation to better evaluate students' essay writing proficiency at the end of a course of study. The study subjects, for example, performed significantly differently from best to least as follows: *Content, Organization, Mechanics, Vocabulary, Language*.

Kim (2005) pointed out four problems in university writing classes: first, *a heavy emphasis on grammatical form*; second, *overemphasis on final product*; third, *lack of genre-specific across the curriculum*; and fourth, *the need for more diverse types of feedback*. To solve these problems, Kim suggested utilizing the balanced instructional and curricular approach of the writing process and the genre – based approach to teaching writing.

Based on these two approaches to teaching, Kim provided four principles (guidelines) that apply to Korean university level writing classes. These four principles (guidelines) are: *balancing form and function, scaffolding language and learning, extending the university curriculum and providing meaningful response and formative assessment*. Uchniat (2005) said that teachers of English compositions should have an agenda, i.e., *goals or objectives to accomplish with students* and apply methods or combinations of methods, traditional or otherwise to do so. Therefore, it seems important that if teachers of English composition periodically review, re-evaluate and then, as a result modify or just confirm their methods in relation to their agenda, they may be able to be as dynamic as they would like their students to be. Warschauer and

Ware (2006: pp. 1-24) indicated that researchers developed online *Automated Writing Evaluation (AWE) programmes* as a way to meet the challenge which has to do with the great amount of time and skill graders/scores needed to evaluate repeated drafts of students' writing. Shokropour and Fallahzadeh (2007) focused on EFL writing problems at the university level, trying to point out the major difficulties with which Iranian medical students face when writing their reports. The specific objective of this study was to determine whether language skills or writing skills are the major problem areas fifth year medical students and interns confront. In order to compare these students, the researchers analyzed 101 admission and progress notes they wrote in the internal medicine and pediatric wards based on systematic sampling approach. They scored the notes for language skills comprising *spelling, vocabulary, grammar and syntax* and writing skills including *punctuation, cohesive devices, coherence and organization*. Data analysis indicated that Iranian EFL medical students have problems in both language and writing skills, but with a higher percentage of problems in writing skills. Shokropour and Fallahzadeh concluded that Iranian EFL medical students need more hours of EFL and writing classes. The researchers also concluded that different approaches to teaching writing cannot be applied in an EFL context successfully unless EFL teachers take their students' social and academic context, needs and purpose of writing into account in their writing classes. Their study also revealed that teachers must emphasize the link between discourse, community and knowledge in an attempt to offer a new insight on EFL writing. Xinhua's (2008) study aimed at examining whether researchers can use *syntactic maturity* as a reliable instrument to investigate the relationship between English speaking and writing. Forty randomly selected college – level ESL students who studied in one American university participated in the study. The subjects' written and spoken samples were their diagnostic essays written for a college – level ESL composition course at one American university. The researchers gathered the subjects' spoken samples through their participation in a semi – direct, tape- mediated oral proficiency test, the Video Oral Communication Instrument (VOCI). The results of the study showed that the same subjects ranked as high – and low – rated groups in both spoken and written samples. That is to say, the subjects who ranked highly in the spoken samples also ranked highly in the written samples and vice versa. In other words, this study demonstrated a positive and significant correlation between spoken and written data – students with good speaking skills had good writing skills and vice versa.

Nguyen and Hudson (2010) examined preservice EFL teachers' attitudes, needs, and experiences about learning to teach writing in English before their practicum in Vietnamese high schools. An open – ended questionnaire called data from 97 preservice EFL teachers at the beginning of their final practicum. The data suggested that these preservice EFL teachers were motivated to learn to teach English in general and teach writing in particular, but required mentors to model effective teaching practices and share their teaching experiences. Generally, these preservice teachers wanted to know how to teach EFL writing with knowledge of different

genres, classroom management techniques, and provide feedback to their students. Classroom issues related to teaching writing such as *writing genres, writing topics, how to motivate students to learn writing and how to deal with mixed – levels of students at secondary schools* need to be incorporated in preservice teacher coursebook. This enhances the developmental processes on learning to teach EFL writing.

To round off the above – mentioned invaluable contributions towards making the process of grading EFL students' compositions reliable, the researchers of the present study argue that further pieces of research on this particular research area are badly needed. The present study, therefore, attempts to make a contribution to this invaluable research area.

Methodology

Purpose of the Study

Proper EFL composition grading is not attained unless EFL Jordanian writing instructors know what to look for in their EFL students' compositions. Therefore, the purpose of the present study is to find out a reliable measure or instrument for EFL Jordanian writing which would increase its grading reliability and improve the quality of its teaching to EFL Jordanian students at the university level. To serve this purpose, the researchers of the present study compared *the analytic composition grading method* with the "*impressionistic*" one in an attempt to answer the following 2 study questions:

1. *Are there significant differences between the analytic composition grading method and the impressionistic one for each of the subjects' writing abilities i.e., Beginning, Intermediate and Advanced?*
2. *Which method has higher inter - rator reliability: the analytic or the impressionistic?*

Procedure

The researchers of the present study first divided the composition task into three main components: *structure, content and organization*. Details of such components along with the questionnaire format, which the researchers gave to 50 experienced and qualified EFL Jordanian teachers with B.A degrees in English as a pilot study, appear below:

Questionnaire Format

1. **Structure** included the grammaticality of it, i.e. its, accuracy: S – V agreement, tense, word order, function words, sentence complexity and variety, accuracy of punctuation, spelling and capitalization.

Running head: ANALYTIC OR IMPRESSIONISTIC EVALUATION OF COMPOSITION

2. **Content** included appropriateness of ideas variety, theme development or logical development of thought.

3. **Organization** included paragraphing, statement of ideas, heading, margins, appropriate length of a paragraph, use of relevant topic sentences and readability / legibility of handwriting.

(Out of 100 Points)			
	Structure	Content	Organization
<p>* Beginning</p> <p>* Intermediate</p> <p>* Advanced</p>			
<p>* "Beginning" is defined as the first year of writing in the EFL programme at the university level,</p> <p>* "Intermediate" is defined as the university level following Beginning, 1 to 2 years after the Beginning level</p> <p>* "Advanced" is defined as the level for the fourth year at university</p>			
<p>Comments:</p> <hr/> <hr/> <hr/> <hr/>			

Most of the questionnaire respondents commented that there was overlapping among these 3 levels, i.e., "Beginning", "Intermediate", and "Advanced". Consequently, the researchers

revised the same questionnaire format, but this time they divided the composition task into 5 main components instead of 3 as follows:

Revised Questionnaire Format

1. **Structure:** The structure component included accuracy, i.e., grammatical accuracy (S – V) agreement, tense, word order, number, function words, etc. sentence complexity and variety of sentence construction.
2. **Content:** The content component included relevance of ideas, elaboration and development of thesis.
3. **Vocabulary:** The vocabulary component included appropriate choice of lexical items, range and register.
4. **Organization:** The organization component included logical theme development plus clear statement of ideas.
5. **Mechanics:** The mechanics component included spelling, capitalization, punctuation, paragraphing and readability / legibility of handwriting.

(Out of 100 Points)					
	Structure	Content	Vocabulary	Organization	Mechanics
* Beginning					
* Intermediate					
* Advanced					
<p>* "Beginning" is defined as the first year of writing in the EFL programme at the university level,</p> <p>* "Intermediate" is defined as the university level following Beginning, 1 to 2 years after the Beginning level</p> <p>* "Advanced" is defined as the level for the fourth year at university</p>					
<p>Comments:</p> <hr/> <hr/> <hr/> <hr/>					

Once again, the researchers gave the revised questionnaire version to another group of 50 experienced and qualified FFL Jordanian teachers with B.A. degrees simply because the previous questionnaire was a pilot study. The researchers asked them to distribute a score of 100 points among the above – mentioned 5 components and do the same for the three different levels, i.e., *Beginning, Intermediate and Advanced* as specified in the questionnaire. The researchers also encouraged these 50 experienced and qualified EFL teachers to give their own suggestions and comments. Based on these respondents' responses, the researchers

developed 3 EFL composition grading scales. They administered 2 tests to evaluate these grading scales. In the first test, the researchers obtained a sample of 30 composition scripts representing 3 different levels of ability, i.e., *Beginning, Intermediate and Advanced* from the EFL programme of Isra Private University. The students were from different educational zones, but they were mostly from *Amman, Zarqa and Madaba*. All students wrote on one composition topic, *"My favourite language"*. The researchers gave the composition scripts to 5 highly experienced instructors of EFL writing, each with acknowledged reputations among their colleagues for their expertise in this area. All had relevant post-graduate degrees in TEFL methodology. The researchers instructed these graders/ scorers to grade these scripts *"impressionistically"* and after one week – due to time limitation – they gave the same 5 graders the other grading scales and asked them to grade the same 30 composition scrips again using these scales. The scorers attached a copy of each grading scale to each composition script according to the level of ability, i.e., *Beginning, Intermediate and Advanced*. The researchers conducted 2 testing procedures in 2 different settings and on 2 different samples on the grading scales. In the first test, the researchers 30 EFL composition scripts covering 3 levels of ability – *Beginning, Intermediate and Advanced* –to 5 graders to grade them *impressionistically*, i.e., *according to their general impression*. The researchers did not inform them of this piece of research design. After one week, the researchers asked the graders to grade the writing papers once again according to the new grading scales. They attached a copy of the appropriate EFL composition grading scale to each script and the researchers asked them were to circle the range of each component – *excellent, good, fair, poor, very poor* – and use their judgment in giving a grade within this range. In the second test of the grading scales, the researchers extended the sample to include 90 composition scripts (all are Jordanian EFL learners) and the period between the two grading sessions to 30 days instead of one week for the sake of testing the reliability of the 3 marking scales, i.e., EFL essay writing marking scales for the 3 subjects' levels: *"Beginning", "Intermediate" and "Advanced"* levels and generalizing the present study findings to those of similar studies on the research area, which could be of great benefit for EFL/ESL teachers/ students in their writing courses / classes. Five EFL teachers doing graduate work in TEFL methodology at Amman Arab University for Graduate Studies, Amman – Jordan graded the 90 papers *impressionistically* and then *analytically*. These 5 judges added a numerical grading scale called *"EFL*

composition scales" was added under the description of each writing component, i.e., the graders judged the merit of each composition component, *Sentence Structure (S)*, *Composition Content (C)*, *Vocabulary Items (V)*, *Composition Organization (O)* and *Writing Mechanics (M)* on a 5 – step continuum from *very poor* to *excellent*. The judges distributed the points for each writing component among the 5 steps on the basis that "*excellent*" ranges from 85 to 100, "*good*" 70-84 , "*fair*" 56 to 70, "*poor*" 31 to 55 "*very poor*" 0 to 30. The 5 judges attached each copy of the grading scale to each writing script when graded. To analyse the present study data, the researchers used the analysis of variance to estimate the reliability for the graders on each of the two composition grading methods, i.e., *analytic method compared with the impressionistic method*. The researchers used this statistical tool because it is likely that the increased reliability for the *analytic composition grading method* is due to the constraining of the number of points to be given to the different components of a composition task (Carroll, 1980).

Results

A. Results related to the first study question:

1: Are there significant differences between the analytic composition grading method and the impressionistic one for each level of the subjects' writing abilities, i.e., Beginning, Intermediate, and Advanced?

The general distribution of responses in the levels of writing ability, i.e., *Beginning, Intermediate and Advanced* appear in Table 1 p. 25. The researchers summed up the range, mean, mode and median as measures of variation – for each composition component in each level of writing ability. These measures appear in Table 2 p. 26. It is worth mentioning that the researchers arrived at the numbers in this Table by calculating the mean and rounding it. For example, they rounded the mean for *Structure* in the *Beginning level* which was 30.1 to 30, for *Content* 15.4 they rounded it to 15, for *Vocabulary* 18.5, they rounded it to 15, for *Organization* 15.1, they rounded it to 15, for *Mechanics* 21.7 they rounded it to 25 because a larger number

of respondents, 12 out of 27 gave it 25 and above. The same procedure applies to the remaining two writing levels, i.e., the *Intermediate* and *Advanced writing levels*.

Table 3, p. 27 is a comprehensive Table which includes the final product for the three writing levels. Once again, all the participating EFL Jordanian teachers in the present study, gave more weight to *Structure* in the *Beginning level* than in *Intermediate and Advanced levels* (30 % , 25 % and 20 % respectively). *They gave content* increasing weights across levels (15 % , 20 % and 30 %), *Vocabulary* represented the most weight in the *Intermediate level* (20 %) and equal weights (15 %) in both the *Beginning and Advanced levels*. *Organization* received increasing attention across the three levels (15 % , 20 % and 25 % respectively, but *Mechanics* received decreasing weights across levels (25% , 15 % and 10 %). The researchers calculated the average range and the average standard deviation to show the difference between the two composition grading methods, i.e., the *analytic* compared with the *impressionistic grading method*. This appears in Table 4, p.28. The *average range* consistently dropped down in the three levels of writing ability, *Beginning, Intermediate and Advanced* when the graders employed the *analytic method*. In other words, the *general mean of the average range* of the three levels dropped 13.5 points when they exploited the *analytic grading method*. The *average standard deviation*, consistently dropped in the three writing levels when the graders also used the *analytic grading method*. The *mean of the average standard deviation* for example, dropped 6.7 points. This indicates that the *analytic grading method* is superior to the *impressionistic one*.

B. Results related to the second question of the study:

Q2: Is the inter –rater reliability of the analytic composition grading method higher than that of the impressionistic one?

The researchers exploited *the analysis of variance* in the second test to estimate the reliability of the judges on the two composition grading methods, i.e., the *analytic method* as well as the *impressionistic method*. The reliability of the judges (r) is based on calculating the value of Θ which is

$$\theta = \frac{\text{Ms between people} - \text{MS within people}}{\text{KMS within people}}$$

* MS = Mean Squares

K = Number of Markers or Judges.

The estimate of the reliability of the mean of K which is 5 in the present study is consequently

$$r_5 = \frac{5\theta}{1 + 5\theta}$$

Table 5, p. 29 displays the reliability estimates for the judges on the two grading methods, i.e., *impressionistic* and *analytic*. The reliability of the *impressionistic grading method* was, unexpectedly, higher than that of *the analytic grading method*.

Discussion

Based on the statistical analysis of the study subjects' responses in Tables 1, 2, and 3 pp. (25-27) , all participating EFL Jordanian teachers tended to give more weight to *Structure* in the *Beginning Level* than in the *Intermediate and Advanced Levels* (30%, 25% and 20% respectively). They gave increasing weights to *content* across the subjects' 3 writing levels (15%, 20% and 30%), *Vocabulary* represented the most weight in the *Intermediate Level* (20%) and equal weights (15%) in both the *Beginning and Advanced Levels*. *Organization* received increasing attention across the 3 writing levels (15% , 20% and 25% respectively), but *mechanics* received decreasing weights across the writing levels (20%, 15% and 10%). This study result brings the researchers to the conclusion that EFL Jordanian students, should master *structure* when they

reach an advanced writing level. EFL Jordanian teachers should give less emphasis to *structure* in the *Beginning Level* to encourage oral communication skills. They should give vocabulary the most emphasis at the *Intermediate*

writing level so that they can focus more on *content* at the *Advanced writing level*. Scorers usually give *writing mechanics* a zero in the *Advanced Level* simply because writing instructors should give them enough attention / weight in both the *Beginning and Intermediate writing levels*. The researchers believe that the average EFL Jordanian learner is assumed to have had some sort of extensive exposure to composition writing in his/her native language, in this case, Arabic and accordingly more emphasis is to be given to *vocabulary and mechanics of writing* rather than to *organization and content* in both the *Beginning and Intermediate writing levels*. Once again, and as it has been said before, teachers usually give less emphasis to *structure* in the *Beginning level* to encourage oral communication skills. The researchers support the division of grades equally among the 5 writing components, i.e., *Structure (S), Content (C), Vocabulary (V), Organization (O) and Mechanics (M)*, because experience has shown that the neglect of overemphasis of any of these writing components at any writing level usually damages present and future EFL writing instruction. The researchers concluded that the study subjects need more hours of EFL and writing classes because they have major problems in both language skills comprising *spelling, vocabulary, grammar and syntax* and composition writing skills including *punctuation, cohesive devices and organization*. To elaborate on this point, writing instructors should pay attention to a number of writing skills: language use (*the ability to write correct sentences*), mechanical skills (*the ability to spell and punctuate*), treatment of content (*the ability to think creatively, develop thoughts and exclude irrelevant information*), stylistic skills (*the ability to manipulate sentences and paragraphs and use sentences effectively*) and judgment skills (*the ability to write in an appropriate manner for a particular purpose with a particular audience in mind*. (Heaton,1988) approved of such invaluable EFL classroom practices. This is on one hand. On the other hand, the researchers of the present study think that there are many other writing elements to be considered in measuring EFL Jordanian students' writing proficiency / performance other than those which they measured in this study. These elements include *handwriting, accuracy, style, diction, relevance, originality, elaboration, layout, coherence, unity and logic*. (Wilkinson, 1989) asserted these writing elements. What is more, EFL Jordanian composition teachers should take their students' social and academic context, needs and purpose of writing into account in their EFL writing classes and emphasize the link between discourse community and knowledge in an attempt to offer a new

insight into EFL writing. Interestingly enough, EFL Jordanian composition teachers should also emphasize the link between English speaking and writing because of the positive and significant correlation between spoken and written data – students with good speaking skills have good writing skills and vice versa. (Shokropour & Fallahzadeh, 2007; Xinhua, 2008) confirmed these

implications for EFL writing instruction. The researchers also concluded that EFL Jordanian writing instructors need to be motivated to learn to teach EFL writing with knowledge of different genres, classroom management techniques; they need to learn to provide their EFL students with immediate or delayed corrective feedback on their errors, they need to learn how to motivate them to learn writing and how to deal with mixed levels of EFL students at both the secondary and university levels. These invaluable EFL writing practices need to be incorporated in preservice EFL Jordanian teacher training programme before EFL Jordanian teachers' practicum in Jordanian schools (Nguyen & Hudson,2010) support these implications for EFL writing instruction. It also seems important for EFL teachers of EFL composition to periodically review, re-evaluate and then, as a result modify or confirm their EFL teaching methods in relation to their agenda, i.e., *their goals or objectives*. This simply means that EFL Jordanian teachers of English composition should identify an agenda for their EFL composition courses particularly at the university level to accomplish with their EFL Jordanian students who are of different language backgrounds, different amounts of EFL writing instruction and different academic status and apply EFL teaching methods or a combination of methods to do so. (Uchniat ,2005) confirms these pedagogical implications for EFL writing instruction.

Table 4 p. 28, shows that the *average range* consistently dropped down in the 3 levels of the subjects' writing ability, i.e., *Beginning, Intermediate and Advanced*, the *general mean of the average range* of these 3 levels of writing ability dropped 13.5 points, the *mean of the average standard deviation* dropped 6.7 points when the graders/scorers used the *analytic composition grading method*. This indicates that the *analytic grading method* was superior to the *impressionistic one*. This study finding is in line with that of (EL-Kilabi's,1996). In the researchers' view, a reason for this superiority is that the *analytic writing grading method* is preferable to the *impressionistic one* because the grades the impressionistic method assigns are less reliable than those the analytic one awards. Another reason for the superiority of the *analytic grading method* over the

impressionistic grading method is that one or more graders usually award a grade based on the total impression of the composition topic as a whole. The graders generally find this method, i.e., the *impressionistic grading method* more tiring and faster than the *analytic grading method*. These justifications are in line with those of (Heaton's ,1982). There is a general agreement among EFL Jordanian writing teachers as to the difficulty of grading EFL Jordanian student composition tasks objectively because of the existence of inconsistent grading, i.e., two EFL writing graders may give two different grades to the same essay or extraordinary differences may appear between the grades of different graders in the grading of an essay. Table 5 p. 29., shows that the inter-rater reliability of the *impressionistic grading method* is higher than that of the *analytic grading method*.

This study finding is in line with that of (Kaczmarek ,1980) who concluded that grader judgments of a subjective sort have substantial reliability and are strongly correlated with similar judgments by independent raters and with objective scores computed over the same compositions. An explanation for this finding could be that, as Carroll (1980) indicated, graders cannot avoid subjectivity in this type of problem, i.e., grading composition because even the *analytic grading method*, of course, is to some degree impressionistic, i.e., when a grader grades a composition impressionistically, he / she, in one way or another, has the analytic categories in mind. What is more, the grade which a grader usually gives to a certain composition impressionistically cannot be rationally defended. On the other hand, the *analytic grading method*, though not different from the *impressionistic grading method* in reliability, definitely has its own invaluable pedagogical advantages over the *impressionistic grading method*. The *analytic composition grading method*, for example, helps the graders of a composition see where their efforts paid off and where they need to direct special attention. They can almost always discover some strength and reason to rejoice and also usually, locate language areas for improvement. Once again, the *inter-rater reliability* of the *impressionistic grading method* was, unexpectedly, higher than that of the *analytic grading method* in the present study. This may be because of the use of essays as a measure of language proficiency which has always been met with opposition because of the existence of *inter – rater reliability*, i.e., two essay graders may give two different grades or (multiple grading) to the same essay. This point represents the problem which belongs to the evaluation of compositions. (Heaton ,1982) supports this study finding. This *inter – rater reliability* of the *impressionistic grading method* which was, unexpectedly,

higher than that of the *analytic method* is because different essay graders often assign the same composition to different grading categories. (Homburg ,1984) supports this study finding. Although the *inter – rater reliability* is a significant problem of the holistic grading of compositions written by EFL students, it appears to be valid as evidenced by its widespread use for the evaluation of EFL students' writing proficiency / performance. (Homburg,1984 & Bacha, 2001) give evidence to this finding. On the other hand, the researchers of the present study believe that careful specification of an analytical writing grading instrument can increase reliability, another term for the objectivity of the *analytic grading method* simply because many segments of composition procedures such as *lexis, syntax, spelling and punctuation* lend themselves to objective testing. This study finding lends support to the arguments put forth by (Brown & Baily,1987; Gannon,1985; Carroll & Hall ,1985). The researchers of the present study think that although there is a general agreement among EFL Jordanian teachers, EFL Jordanian specialists and test designers as to the difficulty of grading an EFL composition objectively, it is quite possible to have a reliable assessment of it. (Carroll & Hall,1985) give evidence to this finding. In the researchers' view, graders/scorers can achieve the objectivity of the *analytic method* as a reliable instrument of grading EFL writing by using objective questions (error –

recognition and error – correction questions) through establishing their relationship to essay writing questions using the *analytic grading method* and then by deciding the relationship between test questions and essay writing questions. Graders can establish the *validity* of the 2 writing tests by means of item analysis and their *reliability* by means of a parallel form test (the Cloze Procedure test). Error – recognition and error – correction questions correlate highly with essay writing questions. (El-Kilabi, 1996) confirmed these pedagogical implications for EFL writing instruction. A survey of the literature shows that many EFL composition graders argue that grading EFL composition still remains a complex process due to the complex nature of the writing skill itself and that it requires time and precision only by highly experienced EFL composition graders.

Conclusion

The present study investigated the possibilities of attaining a proper and reliable EFL composition grading instrument, which is badly needed. Specifically, the purpose of this study was to develop a reliable grading tool for EFL composition which may increase its grading reliability and improve the quality of EFL writing teaching to EFL Jordanian students at the university level. The researchers tested and compared the *analytic grading method* with the *impressionistic one*. The present study findings showed that the analytic method has its pedagogical advantages over the impressionistic one in that it lays the foundation of the relevant elements of good writing and when the writing teachers calculate the marks in the marking scheme for their students they can discover some strength and see where their efforts are successful, where their students need special attention and where to locate areas for improvement. The results of the present study are also of great value to all EFL Jordanian teachers, EFL specialists and EFL test designers who take part in developing and redesigning EFL teacher training programmes particularly their EFL writing training component, in the countries where they teach English as a foreign language. The researchers of the present study recommend that further pieces of research on EFL writing be done to document and analyse the practical pedagogical knowledge of highly experienced EFL Jordanian composition teachers. The researchers believe that documenting how these highly experienced EFL teachers conceptualize their thinking about EFL writing instruction is a necessary step towards developing models of it as well as models of their usual classroom practices for EFL writing assessment in EFL composition courses in Jordanian universities. Once again, this invaluable conceptualization of highly experienced EFL Jordanian composition teachers' thinking about their EFL writing

instruction also provides responsible bodies at the university level with clear rationals for selecting appropriate tasks for written composition assessment and specifying adequate standards for EFL Jordanian students' writing achievement. Statistically, these responsible bodies can further validate the present study findings through investigating several variables that may facilitate the process of EFL composition assessment such as running specialized training workshops to train more and more competent EFL Jordanian composition graders and specifying further categories of proper EFL composition grading scales.

References

- Bacha, N. (2001). *Writing Evaluation: What Can Analytic Versus Holistic Essay Scoring Tell Us?* Libanese American University, Byblos, Lebanon.
- Brown, H.D., & Bailey, K.M. (1987). "A Categorical Instrument for Scoring Second Language Writing Skills". *Language Learning*.
- Gannon, P. (1985). *Assessing Writing: Principles and Practice of Marking Written English*. London: Edward Arnold.
- Carroll, B.J. (1980). *Testing Communicative Performance*, Oxford. Perganon Press Ltd.
- Carroll, B.J., & Hall, P.G. (1985). *Make Your Own Language Tests: A Practical Guide to Writing Language Performance Tests*. Oxford: Pergamon Press
- Cumming, A., (1998). Theoretical Perspectives on Writing. *Annual Review of Applied Linguistics*, 18, 61- 78.
- Cumming, A., & Riazi, A. (2000). Building Models of Second – Language Writing Instruction and Achievement. *Learning and Instruction* 10, 55 -71.
- Cumming, A. (2001). ESL/EFL Instructors' Practices for Writing Assessment: Specific Purposes or General Purposes? *Language Testing* 18, 207 – 224.
- EL-Kilabi, A. (1996). The Aalytic Method: An Objective Procedure for Scoring Essays. *ABHATH AL-YARMOUK: (Lit. & Ling.,* 14 (2), 9-26.
- Farhady, H., & Farzanehnejad, A.R. (1996). An Objective Measure for Evaluating EFL Compositions. *Iranian Journal of Applied Linguistics*, 1, 1, 298-307.
- Hackeling, M. (1991). Error Analysis of English Compositions Written by Japanese Students. *Research Reports of Kochi Medical School Liberal Arts*, 7, 13-28.
- Heaton, J.B. (1988). *Writing English Language Tests*. New York: Longman.

- Heaton, J.B. (1982). *Writing English Language Tests*. Essex: Longman.
- Homburg, A.J. (1984). Holistic Evaluation of ESL Compositions: Can It Be Validated Objectively? *TESOL Quarterly*, 18, 1, 87-103.
- Kaczmarek, C. (1980). "Scoring and Rating Essay Tasks", in J. Oller and K. Perkins (eds.), *Research in Language Testing*, (151- 170). Rowly, Mass: Newbury House Publishers..
- Kammeen, P. T. (1989). Syntactic Skill and ESL Writing Quality. In A. Freedman, I. Pringle, and J. Yalden (eds.), *Learning to Write: First Language / Second Language* (162 -170). New York: Longman, Inc.
- Kim, Y. (2005). *Teaching Korean University Writing Class: Balancing the Process and the Genre Approach*. University of New Mexico, Albuquerque, USA.
- Nguyen, H. T. M., & Hudson, P. (2010). Preservice EFL Teachers' Attitudes, Needs, and Experiences about Teaching Writing and Learning to Teach Writing Before Their Practicum. *Asian EFL Journal*, 12, 3.
- Pennington, M., Costa, V., So, S. Shing, J., Hirose, K., & Niedzielski, K. (1997). The Teaching of English as – a – Second Language Writing in the Asia Pacific Region: A Cross – Country Comparison. *RELC Journal*, 28, 120- 430.
- Shokropour, N.,& Fallahzadeh M. H. (2007). A Survey of the Students' and Interns' EFL Writing Problems in Shiraz University of Medical Sciences. *Asian EFL Journal*, 9, 1.
- Uchniat, J.D. (2005). *Individual Conferencing and EFL Composition Class: Agenda, Method, and Student Reactions*. Oct. National Taiwan University.
- Warschauer, M., & Ware, P. (2006). Automated Writing Evaluation: Defining the Classroom Research Agenda. *Language Teaching Research*, 10, 1-24.
- Wilkinson, A. (1989). Assessing Language Development: The Creditor Project. In A. Freedman, I. Pringle, and Yalden (eds.), *Learning to Write: First Language / Second Language* (67-70). New York: Longman, Inc.

Xinhua, Z. (2008). Is Syntactic Maturity a Reliable Measurement to Investigate the Relationship Between English Speaking and Writing?, *Asian EFL Journal*, 10, 7.

About the authors

Khalaf Falah Al-Makhzoomi is an associate professor in the Department of Curriculum and Instruction, Faculty of Education, Yarmouk University, Irbid-Jordan. He holds a PhD in Teaching English as a Foreign Language (TEFL), and he has a background in English teaching and research into TEFL as well as English for Specific Purposes (ESP). He has taught students and trained teachers of English in Jordan for a long time.

Saleh Mohmoud Farah Freihat is an assistant professor in the English Department, Faculty of Arts, Isra University, Amman – Jordan. He has a PhD in *English Language Curriculum and Instruction*. He has a background of English teaching and research into *Teaching English as a Foreign Language (TEFL)* as well as *English for Specific Purposes (ESP)*. He has taught students and trained teachers in both areas in the United Arab Emirates (UAE), Oman and Jordan.

Appendixes

Table 1: The Frequency and Distribution of the Responses / Grades allotted by the Questionnaire Respondents in the Subjects' 3 Writing Levels.

** NB. Each number indicates the number of EFL Jordanian teachers giving the grades.*

Writing Component	Subjects' Writing Level	Grades Allotted										
		0	5	10	15	20	25	30	35	40	45	50
Sentence Structure	Beginning			1	2	1	5	9	1	7		1
	Intermediate		1	1	6	5	11	1				
	Advanced		1	3	3	11	3	5				
Composition Content	Beginning	1	2	9	3	6	2	3		1		
	Intermediate	1		2	3	11	4	3		1		
	Advanced			3		6	7	5		3		3
Vocabulary Items	Beginning		1	10	3	11	1	1				
	Intermediate			5	4	13	2			1		
	Advanced		1	6	5	12	2					
Composition	Beginning			12	3	9	3					
			1	1	4	11	5	3				

Organization	Intermediate							
				11	8	6		1
	Advanced							
Writing	Beginning	6	1	8	3	8		1
	Intermediate	1	10	5	6	3		
Mechanics	Advanced	1	4	11	5	5		

Table 2: A Summary of the Measures of Variation for Each Writing Component in the Subjects' 3 Writing Levels.

Subjects' Writing Level	Writing Component	Measures of Variation				
		Range	Mode	Median	Mean	Scale
Beginning Writing Level	Sentence Structure	10-15	30	30	30.2	30.1 rounded to 30
	Composition Content	0-40	15	15	16.3	15.4 rounded to 15
	Vocabulary Items	5-30	20	20	15.6	18.5 rounded to 15
	Composition Organization	0-25	15	15	15.4	15.1 rounded to 15
	Writing Mechanics	10-40	20	25	21.7	21.7 rounded to 25
Intermediate	Sentence Structure	10-35	30	30	25.6	25.5 rounded to 25
	Composition	0-40	20	20	20.5	20.1 rounded to 20

Writing Level	Content					
	Vocabulary Items	10-40	20	20	18.3	19.4 rounded to 20
	Composition Organization	2-30	20	20	20.3	20.1 rounded to 20
	Writing Mechanics	5-25	10	15	15	13.3 rounded to 15
Advanced Writing Level	Sentence Structure	5-30	20	20	20.1	20
	Composition Content	10-50	27.4	25	26.6	26.3 rounded to 26
	Vocabulary Items	2-25	20	20	16.4	18.8 rounded to 15
	Composition Organization	20-40	20	25	24.5	23.2 rounded to 25
	Writing Mechanics	0-20	10	10	11.6	10.5 rounded to 10

Table 3: A Summary of the Weights Given to Each Writing Component in Each Writing Level

Writing Component	Subjects' Writing Levels		
	Beginning	Intermediate	Advanced
Sentence Structure	30	25	20
Composition Content	15	20	30
Vocabulary Items	15	20	15
Composition Organization	15	20	25
Writing Mechanics	25	15	10

Table 4: A Summary of the Average Range and the Average Standard Deviation for the Subjects' 3 Writing Levels on the Impressionistic vs. Analytic Composition Grading Methods.

Composition Grading Method	Subjects' Writing Levels			Average Range
	Beginning	Intermediate	Advanced	
Impressionistic	39.1	44.4	39.7	41.0
Analytic	28.4	29.9	24.3	27.3
Difference	10.7	14.5	15.4	13.5
				Average Standard Deviation
Impressionistic	16.0	17.2	16.0	16.4
Analytic	8.2	11.3	9.6	9.7
Difference	7.8	5.9	6.4	6.7

Table 5: The Reliability Estimates for EFL Jordanian Judges on the Methods (Impressionistic vs Analytic Composition Grading Method)

Subjects' Writing Level	Composition Grading Method	
	Impressionistic	Analytic
	Beginning	
Writing Level	.816	.828
Intermediate		
Writing Level	.812	.784
Advanced		
Writing Level	.850	.777
Mean	.826	.796